Taking Fuzzy-Rough Application to Mars: Fuzzy-Rough Feature Selection for Mars Terrain Image Classification

Shang, Changjing; Barnes, David Preston; Shen, Qiang

Published in:
Rough Sets, Fuzzy Sets, Data Mining and Granular Computing

DOI:
10.1007/978-3-642-10646-0_25

Publication date:
2009

Citation for published version (APA):
Shang, C., Barnes, D. P., & Shen, Q. (2009). Taking Fuzzy-Rough Application to Mars: Fuzzy-Rough Feature Selection for Mars Terrain Image Classification. In *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing* (pp. 209-216). Springer Nature. https://doi.org/10.1007/978-3-642-10646-0_25

Taking Fuzzy-Rough Application to Mars*

Fuzzy-Rough Feature Selection for Mars Terrain Image Classification

Changjing Shang, Dave Barnes and Qiang Shen

Dept. of Computer Science, Aberystwyth University, Wales, UK.
{cns,dpb,qqs}@aber.ac.uk

Abstract. This paper presents a novel application of fuzzy-rough set-based feature selection (FRFS) for Mars terrain image classification. The work allows the induction of low-dimensionality feature sets from sample descriptions of feature patterns of a much higher dimensionality. In particular, FRFS is applied in conjunction with multi-layer perceptron and K-nearest neighbor based classifiers. Supported with comparative studies, the paper demonstrates that FRFS helps to enhance the effectiveness and efficiency of conventional classification systems, by minimizing redundant and noisy features. This is of particular significance for on-board image classification in future Mars rover missions.

1 Introduction

The panoramic camera instruments on the Mars Exploration Rovers have acquired a large volume of high-resolution images, which provides substantial information to characterize the Mars environment [1, 4]. Automated analysis of such images has since become an important task, especially for surveying places (e.g. for geologic cues) in Mars [8, 12]. Any progress towards automated detection and recognition of objects within Mars images, including different types of rocks and their surroundings, will make a significant contribution to the accomplishment of this task.

Mars terrain images vary significantly in terms of intensity, scale and rotation, and are blurred with noise. These factors make Mars image classification a challenging problem. One critical step to successfully build an image classifier is to extract and use informative features from given images [3, 7, 9]. To capture the essential characteristics of such images, many features may have to be extracted without explicit prior knowledge of what properties might best represent the underlying scene reflected by the original image. Yet, generating more features increases computational complexity and measurement noise, and not all such features may be useful to perform classification. Thus, it is desirable to employ a technique that can determine the most significant features, based on sample measurements, to simplify the classification process, while ensuring high classification performance.

* Work funded by the Daphne Jackson Trust and the Royal Academy of Engineering.

This paper presents an approach for performing large-scale Mars terrain image classification, by exploiting the recent advances in fuzzy-rough set-based feature selection techniques [6]. As such, fuzzy-rough sets are, for the first time, applied to tasks relevant to space engineering. Experimental results show that this application ensures rapid and accurate learning of classifiers. This is of great importance to on-board image classification in future Mars rover missions. The rest of this paper is organized as follows. Section 2 introduces the Mars terrain images under investigation. Sections 3, 4 and 5 outline the key component techniques used in this work, including feature extraction, (fuzzy-rough) feature selection and feature pattern classification. Section 6 shows the experimental results, supported by comparative studies. The paper is concluded in Section 7.

2 *McMurdo* Panorama Image

This work concentrates on the classification of the *McMurdo* panorama image, which is obtained from the panoramic camera on NASA's Mars Exploration Rover Spirit and presented in approximately true color [4]. Such an image reveals a tremendous amount of detail in part of Spirit's surroundings, including many dark, porous-textured volcanic, brighter and smoother-looking rocks, sand ripple, and gravel (mixture of small stones and sand). Fig. 1 shows the most part of the original *McMurdo* image (of a size 20480×4124). This image, excluding the areas occupied by the instruments and their black shadows, is used for the work here, involving five major image types (i.e. classes) which are of particular interest. These image types are: textured or smoothed dark rock (C1), orange colored bedding rock (C2), light gray rock (C3), sand (C4), and gravel (C5), which are illustrated in Fig. 2. The ultimate task of this research is to detect and recognize these five types of image over a given region.

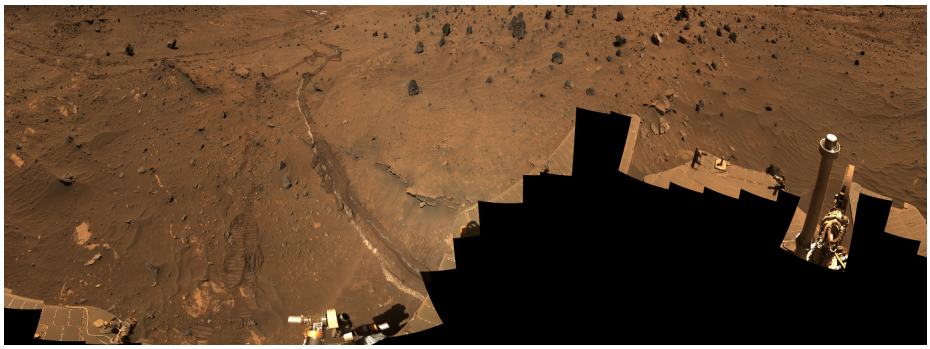

Fig. 1. Mars McMurdo panorama image.

3 Feature Extraction

Many techniques may be used to capture and represent the underlying characteristics of a given image [3, 10]. In this work, local grey level histograms and

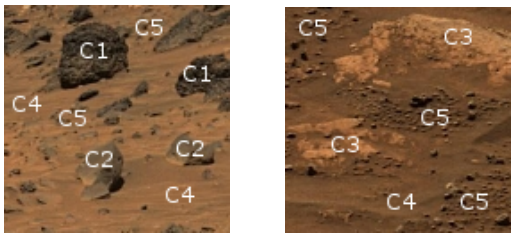

Fig. 2. Image classes (C1: rock1, C2: rock2, C3: rock3, C4: sand, C5: gravel).

the first and second order color statistics are exploited to produce a feature pattern for each individual pixel. This is due to the recognition that such features are effective in depicting the underlying image characteristics and are efficient to compute. Also, the resulting features are robust to image translation and rotation, thereby potentially suitable for classification of Mars images.

3.1 Color Statistics-Based Features

Color images originally given in the RGB (Red, Green and Blue) space are first transformed to those in the HSV (Hue, Saturation and Value) color space [10]. These spaces are in bijection with one another, and the HSV space is widely used in the literature. Six features are then generated per pixel, by computing the first order (mean) and the second order (standard deviation, denoted by STD) color statistics with respect to each of the H, S and V channels, from a neighborhood of the pixel. The size of such neighborhoods is pre-selected by trial and error (which trades off between the computational efficiency in measuring the features and the representative potential of the measured features).

3.2 Local Histogram-Based Features

To reduce computational complexity, in extracting histogram-based features, given color images are first transformed to grey-level (GL) images. For a certain pixel, a set of histogram features $H_i, i = 1, 2, ..., B$, are then calculated within a predefined neighborhood, with respect to a certain bin size B. Here, the neighborhood size is for convenience, set to the same as that used in the above color feature extraction, and H_i denotes the normalized frequency of the GL histogram in bin i. To balance between effectiveness and efficiency, B is empirically set to 16 in this work. In addition, two further GL statistic features are also generated, namely, the mean and STD (which are different from their color statistics-based counterparts of course).

4 Fuzzy-Rough Set-Based Feature Selection

Let U be the set of pixels within a given image, P be a subset of features, and D be the set of possible image classes. The concept of fuzzy-rough dependency measure [6], of D upon P, is defined by

$$\gamma_P(D) = \frac{\sum_{x \in U} \mu_{POS_{R_P}(D)}(x)}{|U|} \tag{1}$$

where

$$\mu_{POS_{R_P}(D)}(x) = \sup_{X \in U/D} \mu_{\underline{R_P}X}(x) \tag{2}$$

$$\mu_{\underline{R_P}X}(x) = \inf_{y \in U} I(\mu_{R_P}(x,y), \mu_X(y)) \tag{3}$$

and U/D denotes the (equivalence class) partition of the image (i.e. pixel set) with respect to D, and I is a fuzzy implicator and T a t-norm. R_P is a fuzzy similarity relation induced by the feature subset P:

$$\mu_{R_P}(x,y) = T_{A \in P}\{\mu_{R_{\{A\}}}(x,y)\} \tag{4}$$

That is, $\mu_{R_{\{A\}}}(x,y)$ is the degree to which pixels x and y are deemed similar with regard to feature A. It may be defined in many ways, but in this work, the following commonly used similarity relation [5] is adopted:

$$\mu_{R_{\{A\}}}(x,y) = 1 - \frac{|A(x) - A(y)|}{A_{max} - A_{min}} \tag{5}$$

where $A(x)$ and $A(y)$ stand for the value of feature $A \in P$ of pixel x and that of y, respectively, and A_{max} and A_{min} are the maximum and minimum values of feature A. The fuzzy-rough set-based feature selection (FRFS) method works by greedy hill-climbing. It employs the above dependency measure to choose which features to add to the subset of the current best features and terminates when the addition of any remaining feature does not increase the dependency.

5 Image Classifiers

Multi-layer perceptron neural networks [11] and K-nearest neighbors (KNN) [2] are used here to accomplish image classification, by mapping input feature patterns onto the underlying image class labels. For learning such classifiers, a set of training data is selected from the typical parts (see Fig. 2) of the *McMurdo* image, with each pixel represented by a feature pattern which is manually assigned an underlying class label.

6 Experimental Results

From the *McMurdo* image of Fig. 1, a set of 270 subdivided non-overlap images with a size of 512×512 each are used to perform this experiment. 816 pixel points are selected from 28 of them, which are each labeled with an identified class index (i.e. one of the five image types: rock1, rock2, rock3, sand and gravel) for training and verification. The rest of all these images are used as unseen data for classification. Each training pixel is represented by a pattern of 24 features (see Section 3). Of course, the actual classification process only uses

Table 1. Feature meaning and reference

No.	Meaning	No.	Meaning	No.	Meaning	No.	Meaning	No.	Meaning
1	Mean(GL)	2	STD(GL)	3	Mean(H)	4	STD(H)	5	Mean(S)
6	STD(S)	7	Mean(V)	8	STD(V)	9-24	H_i		

subsets of selected features. The performance of each classifier is measured using classification accuracy, with ten-fold cross validation.

For easy cross-referencing, Table 1 lists the reference numbers of the original features that may be extracted, where $i = 1, 2, ..., 16$. In the following, for KNN classification, the results are first obtained with K set to 1, 3, 5, 8, and 10. For the MLP classifiers, to limit simulation cost, only those of one hidden layer are considered here with the number of hidden nodes set to 8, 12, 16, 20, or 24. Those classifiers which have the highest accuracy, with respect to a given feature pattern dimensionality and a certain number of nearest neighbors or hidden nodes, are then taken for performance comparison.

6.1 Comparison with the Use of All Original Features

This subsection shows that, at least, the use of a selected subset of features does not significantly reduce the classification accuracy as compared to the use of the full set of original features. For this problem, FRFS returns 8 features, namely, STD(GL), Mean(H), STD(H), Mean(S), STD(S), Mean(V), H_4, H_{15} (i.e. features $2, 3, 4, 5, 6, 7, 12$ and 23 in Table 1), out of the original twenty-four. That is, a reduction rate of two-third. Table 2 lists the correct classification rates produced by the MLP and KNN classifiers with 10-fold-cross-validation, where the number (N) of hidden nodes and that (K) of the nearest neighbors used by these MLP and KNN classifiers are also provided (in the first column).

Table 2. FRFS-selected vs. full set of original features

Classifier	Set	Dim.	Feature No	Rate
MLP(N=20)	FRFS	8	$2, 3, 4, 5, 6, 7, 12, 23$	94.0%
MLP(N=20)	Full	24	$1, 2, ..., 23, 24$	94.0%
KNN(K=8)	FRFS	9	$2, 3, 4, 5, 6, 7, 12, 23$	89.1%
KNN(K=5)	Full	24	$1, 2, ..., 23, 24$	89.2%

The results demonstrate that the classification accuracy of using the eight FRFS-selected features is the same as that of using the twenty-four original features for MLP classifiers (94.0%), and is very close to that for KNN classifiers (89.1% vs. 89.2%). This is indicative of the potential of FRFS in reducing not

only redundant feature measurements but also the noise associated with such measurements. Clearly, the use of FRFS helps to improve both effectiveness and efficiency of the classification process. Note that although the number of original features is not large, for on-board Martian application, especially in relation to the task of classifying large-scale images, any reduction of feature measurements is of great practical significance.

6.2 Comparison with the Use of PCA-Returned Features

Principal component analysis (PCA) [2] is arguably one of the most popular methods for dimensionality reduction, it is adopted here as the benchmark for comparison. Fig. 3 shows the classification results of the KNN and MLP classifiers using a different number of principal features. For easy comparison, the results of the KNN and MLP that use 8 FRFS-selected features are also included in the figure, which are represented by $*$ and $\circ$, respectively.

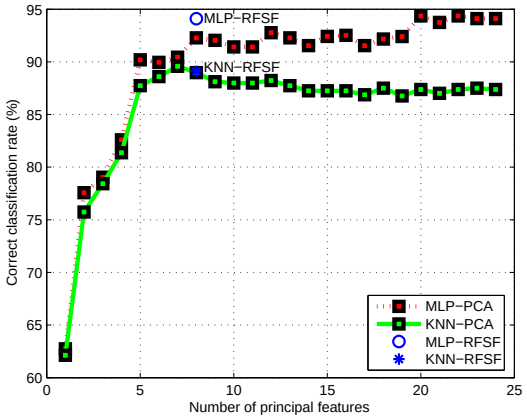

Fig. 3. Performance of KNN and MLP vs. the number of principal components.

These results show that the MLP classifier which uses FRFS-selected features has a substantially higher classification accuracy amongst all those classifiers using a subset of features of the same dimensionality (i.e. 8). This is achieved via a considerably simpler computation, due to the substantial reduction of the complexity in input patterns. The results also show the cases where PCA-aided (MLP or KNN) classifiers each employ a feature subset of a different dimensionality. However, these classifiers still generally underperform than the corresponding FRFS-aided ones, whether they are implemented using MLP or KNN. This situation only changes when almost the full set of PCA-returned features is used where the MLP classifiers may perform similarly or slightly better (if 20 or 22 principal components are used). Yet, this is at the expense of requiring many more feature measurements and much more complex classifier structures. Besides, PCA-returned features lose the underlying meaning of the original.

6.3 Classified and Segmented Images

The ultimate task of this research is to classify Mars panoramic camera images and to detect different objects or regions in such images. The MLP which employs the 8 FRFS-selected features, and which was trained by the given 816 labeled feature patterns, is taken to accomplish this task: the classification of the entire image of Fig. 1 (excluding those regions as indicated previously). As an illustration, three classified images are shown in Fig. 4, numbered by (a), (b) and (c) respectively, where five different colors represent the five image types (rock1, rock2, rock3, sand and gravel). From this, boundaries between different class regions can be identified and marked with white lines, resulting in the segmented images also given in Fig. 4, numbered by (d), (e) and (f) correspondingly.

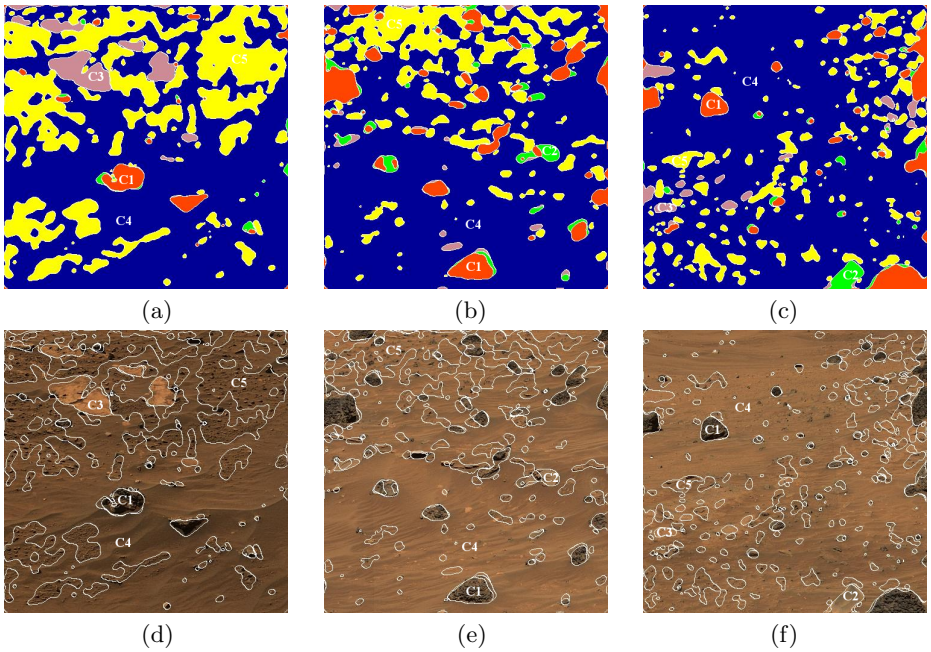

Fig. 4. Classified and segmented image.

From these classified images, it can be seen that the five image types vary in terms of their size, rotation, color, contrast, shapes, and texture. For human eyes it can be difficult to identify boundaries between certain image regions, such as those between sand and gravel, and those between rock2 and sand. However, the classifier is able to perform under such circumstances, showing its robustness to image variations. This indicates that the small subset of features selected by FRFS indeed convey the most useful information of the original. Note that classification errors mainly occur within regions representing sand and gravel. This may be expected since gravel is itself a mixture of sand and small stones.

7 Conclusion

This paper has presented a study on Mars terrain image classification, supported by advanced fuzzy-rough set-based feature selection techniques. For the first time, fuzzy-rough sets have been adopted to help solving problems in space engineering. Although the real-world images encountered are large-scale and complex, the resulting feature pattern dimensionality of selected features is manageable. Conventional classifiers such as MLP and KNN that are built using such selected features generally outperform those using more features or an equal number of features obtained by classical approaches represented by PCA. This is confirmed by systematic experimental investigations (though the influence of parameter set-up for feature extraction, e.g. the number of pixels in neighbors and that of bins in histograms, requires further investigation). The work helps to accomplish challenging image classification tasks effectively and efficiently. This is of particular significance for classification and analysis of real images on board in future Mars rover missions.

References

1. Castano R. *et al.*: Current results from a rover science data analysis system. Proc. of IEEE Aerospace Conf. (2006)
2. Duda R. O., Hart P. E., Stork D. G.: Pattern classification. (2nd edition). Wiley & Sons, New York (2001)
3. Huang K., Aviyente S.: Wavelet feature selection for image classification. IEEE Trans. Image Proc. 17, 1709-1720 (2008)
4. http://marswatch.astro.cornell.edu/pancam_instrument/mcmurdo_v2.html
5. Jensen R., Shen Q.: New approaches to fuzzy-rough feature selection. IEEE Trans. Fuzzy Syst., 17(4):824-838 (2009)
6. Jensen R., Shen Q.: Computational intelligence and feature selection: rough and fuzzy approaches. IEEE Press and Wiley. (2008)
7. Kachanubal T., Udomhunsakul S.: Rock textures classification based on textural and spectral features. Proc. of World Academy of Science, Eng. and Tech. 29,110-116 (2008)
8. Kim W. S., Steele R. D., Ansar A. I., Al K., Nesnas I. Rover-Based visual target tracking validation and mission infusion. AIAA Space. 6717-6735 (2005)
9. Lepisto L., Kunttu I., Visa A.: Rock image classification based on k-nearest neighbour voting. Vis. Im. and Sig. Proc., IEE Proc. 153(4), 475-482 (2006)
10. Martin D. R., Fowlkes C. C., Malik J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans. Patt. Anal. and Mach. Inte. 26, 530-549 (2004)
11. Rumelhart D., Hinton E., Williams R.: Learning internal representations by error propagating. In: Rumelhart D., McClell J. and (Eds.): Parallel Distributed Processing. MIT Press. (1986)
12. Thompson D. R., Castano R.: Performance comparison of rock detection algorithms for autonomous planetary geology. Proc. of IEEE Aerospace Conf. paper no. 1251 (2007)